

Exercise 11 - Bioinformatics

Complete the tasks below and submit your results via a pull request on GitHub by the beginning of tutorial next Friday.

To begin this week, fork the TA's Exercise 11 Github repo. Clone the forked repo so that you have the required files. Be sure to commit regularly to show how you arrived at your solutions.

1. You have been provided with eight **.fasta** files containing gene sequences. Four of these have "sporecoat" in the file name and contain sequences for a gene involved in creating bacterial spores. The other four files have "transporter" in the file name and contain sequences for a gene involved in sugar transport into the cell by bacteria. From these files, generate two sequence alignments - one for the spore formation gene and one for the transporter gene. Store your code in a Unix shell script.
2. Write a shell script that generates a HMM profile for the transporter gene and uses HMMer to search the eight bacterial proteomes provided as fasta files. Your script should record the file name and number of hits to the HMM profile in each proteome in a single text file.

Turning in your assignment via GitHub

Once you have committed all changes to your local Git repo and pushed all of those commits to the forked repo on GitHub, you can "turn in" your assignment using a **pull request**. This can be done from the GitHub repo website. When viewing the forked repo, select "Pull requests" in the upper middle of the screen, then click the green "New pull request" button in the upper right. You'll then see a screen with a history of commits, select the green "Create pull request button". In the text box next to your user icon near the top of the page, remove whatever text is there and add "last name submission", but obviously substitute your last names. Then click the green "Create pull request" button.